

# Physicochemical characterization and functional site analysis of lactoferrin gene of Vechur cow

Anisha Shashidharan, Ruchi Singh, Salini Bhasker, Chinnamma Mohankumar\*

SCMS Institute of Bioscience and Biotechnology Research and Development, South Kalamassery, Cochin- 682033, India; Chinnamma Mohankumar - Email: mohankumar@scmsgroup.org; \*Corresponding author

Received May 18, 2011; Accepted June 01, 2011; Published June 23, 2011

## Abstract:

The *Bos indicus* Vechur breed cow milk is known for its medicinal value and the breed is listed under the category of critically maintained breeds by the Food and Agriculture Organization. The lactoferrin protein in milk is known for its nutritional value. Gene polymorphisms have been reported for Bovine lactoferrin. Mutations in the evolutionarily conserved sites tend to impair protein function and are related with the physicochemical difference between the known variants with 11 SNPs within the wild type. Structural differences are located due to these SNPs that may lead to functional variations. The structural variation is seen primarily in the first 48 residues at 5' end in all the samples modelled. Out of 11 SNPs 5 amino acid variations fall under alpha helix and beta sheet region, this might be of functional significance. This result may provide evidence that the SNPs detected in lactoferrin gene might have potential effects on milk composition. Our result demonstrates one major domain that could be a common binding pocket to all the samples, and important as an active site common to all the breeds that could be utilized for effective drug designing. Moreover, at some SNP positions in Vechur breed, antimicrobial peptides were located indicating importance of those residues for enhanced antimicrobial activity in lactoferrin of Vechur breed. Second binding pocket found in N- lobe region with the three required residues aspartic acid, histidine and tyrosine for iron binding, was considered as major binding site.

**Keywords:** Vechur; Critical Breed; Lactoferrin; Bioinformatics; SNP.

## Background:

India possesses the largest population of cattle in the world and Vechur (*Bos indicus*) is one of the indigenous cattle breeds of Kerala, India. It is the smallest cow (nearly 90 cm high) in the world that produces relatively high amount of milk as compared to other breeds. Moreover, it has got a low level of feed requirements and high disease resistance [1]. Milk of Vechur cow has got a high proportion of small fat globules and saturated fatty acids, hence suitable for infants and sick [1]. This milk is considered of medicinal value and is used in Ayurveda system of medicine. Food and Agriculture Organization (FAO) has listed Vechur cow breed under the category of critical breeds. Lactoferrin (LTF) belongs to the serum transferrin gene family and the bovine LTF gene is localized to chromosome 22 and syntenic group U12 [2]. It is a potential gene for milk composition and body measurement traits in dairy cows. Though, genotype, sex, age, breed, herd, location, and other random environmental factors affect milk composition in the organism, single nucleotide polymorphisms (SNPs) reportedly have some very important role. Though most of the SNPs are silent, without an observable phenotype, some SNPs make an individual unique, genetically. Apart from LTF's antibacterial and tumor growth-inhibiting functions, its physiological roles in bone growth and the efficient expression of the milk protein are also very important [3]. Few SNPs of the bovine LTF gene have been reported. Associations between bovine LTF gene polymorphism and somatic cell count in milk produce are well reported [4] by many scientists but are not much studied in this species. The information required for determining and characterizing a protein's function are hidden in its amino acid sequence. The physicochemical properties of the amino acids present in the proteins are well understood with the use of

computational tools used in proteomics and molecular biology [5]. If the bioinformatics methods are standardized and optimized, it may lead to the characterization of sequences and identification of functional targets. Moreover, many attempts have been well documented for characterization and prediction of functional pockets of a query protein by its amino acid sequence [6]. It is widely recognized that mutations in evolutionarily conserved sites tend to impair protein function. There is also weak evidence that protein impairment are somewhat correlated with the physicochemical difference between the original amino acid and the missense variant [7]. Much attention has been focused on the possible phenotypic effect of SNPs that cause amino acid changes. The volume of available information together with the development of more sophisticated methods of protein structure prediction has led to different attempts of relating the effect of amino acid changes to structural distortions and, consequently possible phenotypic effects [8]. Recently we cloned and sequenced the full coding region of LTF gene of Vechur breed of *B. indicus*. The objectives of the present study are to detect SNPs in the LTF gene, its physicochemical characterization and binding site analysis in some animals of *B. indicus* Vechur breed and its comparison with 'wild type' *B. taurus* using bioinformatics tools.

## Materials and Methodology:

### Analysis and comparison of amino acid sequences:

MUSCLE (Multiple Sequence Comparison by Log- Expectation) software employed for comparative analysis, exercises a progressive algorithm based method that not only provides better average accuracy but better speed and hence was employed for multiple sequence alignment [9]. The translated

sequences were then again compared to check if these variations cause substitution of one amino acid for another in the protein (missense mutation). Physicochemical analysis of the sequences namely molecular weight, theoretical pI, amino acid composition, instability index, aliphatic index and grand average of hydropathicity (GRAVY) were performed using PROTPARAM tool [http://www.expasy.ch/tools/protparam.html]. Antimicrobial peptide prediction was computed using APD2: Antimicrobial Peptide Predictor [aps.ummc.edu/AP/about.php] and AntiBP Server [www.imtech.res.in/raghava/antibp/index.html].

### Structure prediction and analysis:

LTF protein homologs from Vechur breed and the *B. taurus* sequences were modelled using 'Modeller9v8' software which implements comparative protein structure modelling by spatial restraints to construct protein models [10, 11]. LTF protein variants were modelled using structure of 'diferric buffalo lactoferrin' (PDB ID: 1BIY) obtained from protein data bank. To gain more insight into locations of these amino acids, the domains were predicted and analyzed using PDBsum from Procheck. Procheck verifies the stereo chemical quality of a protein structure and provides an at-a-glance overview of every macromolecular structure deposited in the Protein Data Bank. PyMOL program was employed for interactive visualization and analysis of molecular structures [12]. Structural superposition was performed between the *B. indicus* protein variants and the *B. taurus* protein sequence using a combination of MatchMaker and Match- Align subroutine of Chimera [13]. A novel score incorporating both secondary structure and residue type is used to align the sequences in Match Maker while Match- Align constructs sequence alignments from pre-existing superpositions of structures. These tools when used together enhance the understanding of sequence information in the context of structure and vice versa [14].

### Binding pocket exploration:

Surface topography of proteins was computed using CASTp server [15]. The weighted Delaunay triangulation and the alpha complex for shape measurements method is utilized by CASTp server that compute identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules. This tool measures analytically the area and volume of each pocket and cavity, both in solvent accessible surface and molecular surface.

### Result and Discussion:

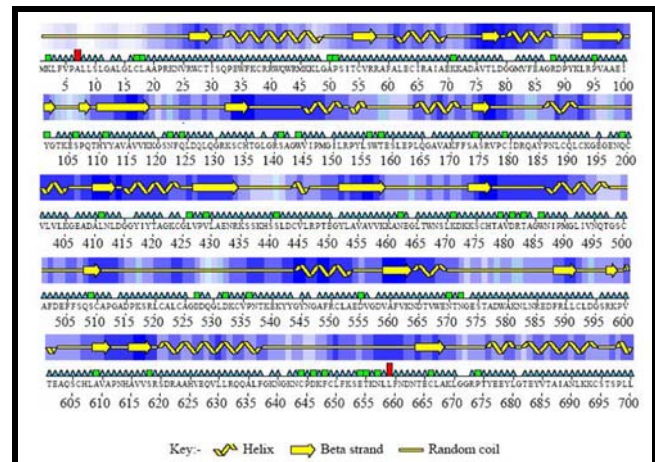
#### Comparative sequence analysis:

In the present study, the full coding region of the LTF gene of Vechur breed of *B. indicus* was cloned and sequenced. The LTF sequences of five animals of *B. indicus* Vechur breed namely LTFV1 (NCBI ID: ACY01187), LTFV2, LTFV3, LTFV4 and LTFV5 (Accession Nos. JF926526, JF926527 and JF926528) were analyzed using various bioinformatics tools to find variation in LTF of Vechur breed and *B. taurus* LTF represented as LTFt (NCBI ID: AAA30609). These variations in the sequences could be the probable cause for better milk yield in Vechur breed. Comparative sequence analysis for Vechur breeds LTF gene samples with 'wild type' *B. taurus* revealed fifteen single nucleotide variations at the very onset using MUSCLE [9] software, resulting in eleven amino acid differences. The assembled results present evidence that, the difference in LTF gene might have potential effect on milk composition in these species. Arginine residue at position 176 in all five LTFV samples, at 264 in LTFV3 and LTFV4 while at 632 position in LTFV3, LTFV4 and LTFV5 are of prime importance. This study explains primarily the effect of SNPs detected in the LTF gene. The nonsynonymous mutation that results in more arginine residue in LTFV samples further illustrate powerful bioactivity of LTF in the Vechur breed that is expected to have an increase of a part of phenotypic variation, especially on milk protein.

#### Physicochemical characterization:

Table 1 (see Supplementary material) provides details of the physicochemical analysis which shows more of the positively charged residues in the sequences and shows instability index ranging from 40.47 to 41.34. A protein whose instability index is smaller than 40 (+5%) is predicted as stable hence these proteins might be stable in a test tube. GRAVY indicates the solubility of the proteins and was computed negative in all the samples (hydrophilic). Calculation of protein isoelectric point (theoretical pI) and molecular weight (Mw) of a specified amino acid sequence are useful to know the approximate region of a 2-D gel where a protein may be found. Theoretical pI ranges from 8.69 to 8.76 in the given samples, showing protein pI as basic. The PROTPARAM tool calculates an accurate (+/- 5% in most cases) molar extinction coefficient for proteins at 280 nm, simply from knowledge of the

amino acid composition. The extinction coefficient indicates how much light a protein absorbs at a certain wavelength and is useful while purifying a protein. Our results establish that difference in sequence at discovered positions leads to physicochemical variation and is highly predictive of the impact of amino acid substitutions on protein function. Antimicrobial peptides are an evolutionarily conserved component of the innate immune response which demonstrates it as novel therapeutic agents. In our desire to better understand the functions of antimicrobial properties of LTF in innate immunity in Vechur breed with that of the wild type, APD2: Antimicrobial Peptide Predictor and AntiBP Server was employed for the prediction of the antibacterial peptides using antimicrobial peptide database (APD). The distribution of sequences in antimicrobial motif was predominantly studied that result in many potential antibacterial peptides that need to be investigated further for confirmation. These softwares confidently predict antimicrobial peptides at 61 and 428 positions and show some chance of being an antimicrobial peptide at 538 and 546 positions in the LTFV samples which was not found in LTFt. This could be the cause for enhanced innate immunity in Vechur breed (Table 2 see Supplementary material). Functional difference in Vechur breeds and the *B. taurus* LTF may be due to the differences in standard physicochemical properties.

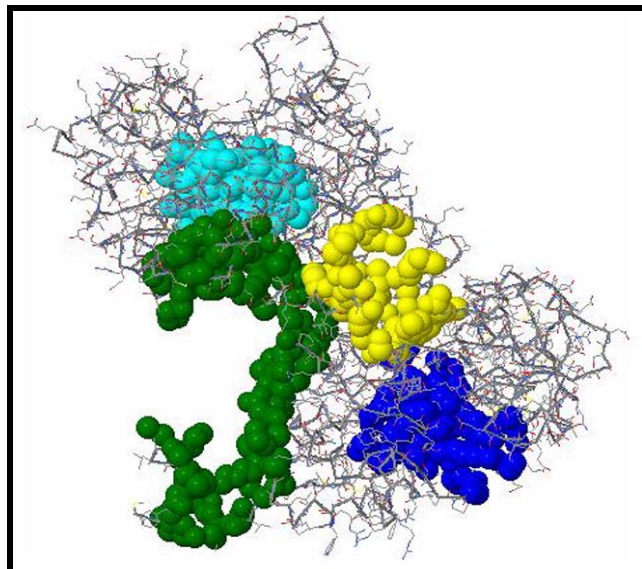


**Figure 1:** A curtailed output of secondary structure elements in LTFV5: Alpha helix, beta sheet and random coil location in the curtailed output from residue property plot. The eleven amino acid residue variations have revealed a beta sheet region at 176 and 428 positions, helix region at 546, 627 and 632 positions while others are found to be in random coil region.

#### Structure prediction and analysis:

In order to determine difference in tertiary structure LTFV protein variants and the reference LTFt samples were modeled using coordinates of diferric buffalo LTF (PDB: 1BIY) [16] and the modeled structures were validated with the Ramachandran plot as well as its RMSD value. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% residues in the most favored regions [17]. LTFV1 and LTFV2 resulted 90.5%, LTFV3 and LTFV4 resulted 90.6% while LTFV5 resulted 91.5% residues in the most favored regions in the Ramachandran plot, hence was designated as a stable structure. The RMSD value also scored less than one, an acceptable range for all the samples modeled. To gain more insight into locations of these amino acids, the modeled structures were superimposed using Match- Align subroutine of Chimera software. The result demonstrates moderate difference in structure at 5' end for 48 residues. This could be quite significant as residue property plot confirmed residues ranging from 30 to 48 forming helical region (Figure 1). Moreover, the structure analysis from residue property plot confirmed the secondary structure elements of some of the SNP positions in beta sheet region (176 and 428 positions) and helix region (546, 627 and 632 positions) (Figure 1) indicating an important structural and functional role of these residues. It is very interesting to note that LTFV samples have more alpha helical strands than beta sheets which are also found more than LTFt sample when compared. This can be of functional significance especially with the disparity in some of the alpha helices. This finding is substantial and falls in line with the previously documented studies of computational protein design efforts on the surface of GCN4, a homodimeric coiled coil that suggest helix propensity is the key factor in sequence design for surface helical positions [18]. The difference in size of the side chains of residues suggests that it may

have significant local influence on the structure. Result from SMART genomes [19] server suggests three transferrin domains in LTFV samples, one low complexity motif from residue 7 to 20, and two confidently predicted domains from 25 to 363 positions and from 364 to 697 positions.



**Figure 2:** Binding pocket location LTFV1: The figure from CASTp software shows a wire frame structure of LTFV1, showing four major domains in different-colored space fill model.

### Binding pocket exploration:

While many surface sites in proteins can tolerate a wide variety of side chains, surface residues can still have a significant effect on protein structure and stability. Binding sites and active sites of proteins and DNAs are often associated with structural pockets and cavities. The binding pocket search resulted in 136 binding pockets that did not show reasonable difference in all the structures except that of LTFV5, and need to be investigated further. This result suggests that despite of some variation leading to physicochemical variations, the binding sites in four LTFV samples are exactly similar as in LTFBt, while it is found rather different in LTFV5 sample. Based on area and volume of the binding pockets, four pockets were selected as major binding pockets for further study (Table 3 see Supplementary material). This study has its importance in drug development based on LTF protein structure, as one drug developed for wild type can also be employed for other variants with similar structure. The procured consensus region may be utilized for effective drug designing [20]. Subsequently, analysis was performed to discover binding pocket position and the residues located in it (Table 3). Binding pocket exploration produced a total of 136 binding pockets. Four binding pockets were considered 'major' based on area and volume of the site. The largest binding pocket (Green color in Figure 2) discloses eight aspartic acid residues, one tyrosine residue and no histidine residue. The second (Blue color in Figure 2) and fourth (Yellow color in Figure 2) binding pockets show all the three amino

acid residues but one aspartic acid. Third pocket though (Cyan color in Figure 2) provides all the three amino acid residues in the required number for iron binding that is two aspartic acids and one each of histidine and tyrosine. The 708 amino acids of LTF, forms two homologous globular domains named N- and C-lobes. N-lobe corresponds to amino acid residues 1-333 and C-lobe to 345-692. Literature suggests that N lobe is important for binding and the ends of the domains are connected by a short  $\alpha$ -helix [21]. The metal binding sites of transferrin domains are localized in each of the two protein globules. There, each ion is bonded with six ligands: four from the polypeptide chain (two tyrosine residues, one histidine residue and one aspartic acid residue) and two from carbonate or bicarbonate ions. Therefore, we can establish as per the literature discussed above that the second binding pocket is considered more prominent for binding, as it does not only contain all the three required residues for iron binding but also falls in N-lobe region.

### Conclusions:

In conclusion, SNPs identified in the LTFV samples result in missense mutations as well. The sequence and structural difference due to these variations was investigated in the present research. Our results provide evidence that the SNPs detected in the LTF gene might have potential effects on milk composition. Binding site analysis demonstrates one major domain that could be a common binding pocket to all the samples that could be utilized for effective drug designing. Furthermore, antimicrobial peptide regions at SNP positions might be a reason for improved antimicrobial property of the milk of Veurcher breed. Therefore, further study will be essential using the SNP for marker-assisted selection in larger populations to substantiate the result.

### Acknowledgements:

We are thankful to Chairman, SCMS Group of Educational Institutions for providing infrastructure facilities for carrying out the work.

### References:

- [1] Lali FA & Bindu KA. *Veterinarski Arhiv*. 2011 **81**(1): 35
- [2] Schwerin M *et al. Mamm Genome*. 1994 **5**: 486 [PMID: 7949732]
- [3] Adlerova *et al. Veterinarni Medicina*. 2008 **53** (9): 457
- [4] Wojdak-Maksymiec K *et al. Veterinarni Medicina*. 2006 **51**(1): 14
- [5] Andrade MA & Sandrer C. *Curr Opin Biotechnol*. 1997 **8**: 675
- [6] Sharma R *et al. Bioinformatics* 2009 **4**(1): 12 [PMID: 20011147]
- [7] Stone EA & Sidow A. *Genome Res*. 2005 **15** (7): 978 [PMID: 15965030]
- [8] Conde L *et al. Nucleic Acids Res*. 2004 **32**(2): W242 [PMID: 15215388]
- [9] Edgar RC. *BMC Bioinformatics*. 2004 **5**: 113 [PMID: 15318951]
- [10] Sali A & Blundell TL. *J Mol Biol*. 1993 **234** (3): 779 [PMID: 8254673]
- [11] Fiser A *et al. Protein Sci*. 2000 **9** (9): 1753 [PMID: 11045621]
- [12] www.pymol.org
- [13] Pettersen EF *et al. J Comput Chem*. 2004 **25**(13):1605 [PMID: 15264254]
- [14] Meng EC *et al. BMC Bioinformatics*. 2006 **7**: 339 [PMID: 16836757]
- [15] Dundas J *et al. Nucleic Acid Res*. 2006 **34**: W116 [PMID: 16844972]
- [16] Berman HM *et al. Nucleic Acids Res*. 2000 **28**(1): 235 [PMID: 10592235]
- [17] Laskowski RA *et al. J Appl Cryst*. 1993 **26**(2): 283
- [18] Pavel Strop *et al. Protein Sci*. 2000 **9**: 1391 [PMID: 10933505]
- [19] Schultz J *et al. Proc Natl Acad Sci U S A*. 1998 **95**: 5857 [PMID: 9600884]
- [20] Sharma R & Singh AK. *Am J Infectious Diseases*. 2009 **5**: 149
- [21] Kuwata H *et al. Biochem Biophys Res Commun*. 1998 **245**(3): 764 [PMID: 9588189]

Edited by P Kanguane

Citation: Shashidharan *et al. Bioinformatics* 6(7): 275-278 (2011)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

## Supplementary material:

**Table 1:** Physicochemical properties of LTf sequences

| S. No. | Sample | MW      | NCR | PCR | TP   | II    | AI    | G      |
|--------|--------|---------|-----|-----|------|-------|-------|--------|
| 1      | LTfBt  | 78056.3 | 76  | 92  | 8.69 | 40.99 | 78.16 | -0.289 |
| 2      | LTfV1  | 78109.4 | 76  | 93  | 8.74 | 40.91 | 78.16 | -0.299 |
| 3      | LTfV2  | 78109.4 | 76  | 93  | 8.74 | 40.91 | 78.16 | -0.299 |
| 4      | LTfV3  | 78186.4 | 77  | 94  | 8.76 | 40.47 | 78.43 | -0.306 |
| 5      | LTfV4  | 78186.4 | 77  | 94  | 8.76 | 40.47 | 78.43 | -0.306 |
| 6      | LTfV5  | 78158.4 | 77  | 93  | 8.70 | 41.34 | 77.88 | -0.308 |

SN = S. No.; MW = molecular weight; NCR = negatively charged residues; PCR = positively charged residues; TP = Theoretical pI; II = Instability index; AI = Aliphatic index; G = Gravity

**Table 2:** Antimicrobial peptide table: Showing antimicrobial peptides computed at SNP positions in LTfV sample

|     |   |
|-----|---|
| 61  | KLGAPSITCVRRFAV                                       |
| 428 | GYIYTAGKCGLVSVL<br>YIYTAGKCGLVSVLA<br>GLVSVLAENRKSSKH |
| 538 | CAGDDQGLDKCVPNT<br>GDDQGLDKCVPNTKE                    |
| 546 | YNGAFRCLAEDVGDV<br>GYNGAFRCLAEDVGD                    |

**Table 3:** Binding pocket exploration: 136 pockets were identified. Based on area and volume of the site four pockets were discovered as major binding pockets. The largest binding pocket (Green color in figure 3) discloses eight aspartic acid residues, one tyrosine residue and no histidine residue. The second (Blue) and fourth (yellow) binding pockets show all the three amino acid residues but aspartic acid is only one. The third pocket provides all the three amino acid residues in the required number for iron binding that is two aspartic acid and one each of histidine and tyrosine.

| Pocket | Area<br>A <sup>2</sup> | Volume<br>A <sup>3</sup> | Color  | Amino acid residues (Name: Position)  |
|--------|------------------------|--------------------------|--------|---|
| 136    | 1844.7                 | 5656.1                   | Green  | K:2 L:3 F:4 V:5 P:6 A:7 L:9 S:10 L:11 A:13 L:14 G:15 L:16 C:17 L:18 A:19 A:20 P:21 R:22 K:23 A:50 P:51 S:52 K:92 R:94 R:277 S:278 V:279 <b>D:280</b> G:281 K:282 E:283 <b>D:284</b> L:285 K:288 R:328 P:330 S:331 K:332 <b>D:334</b> L:337 S:394 T:395 T:396 <b>D:397</b> <b>D:398</b> V:401 K:405 <b>Y:417</b> T:420 V:480 <b>D:481</b> W:486 N:487 F:502 <b>D:503</b> G:513 A:514 <b>D:515</b> P:516 K:517 S:518 L:520 <b>D:532</b> K:533 C:534 V:535 P:536 L:691 K:692 C:694 S:695 T:696 S:697 P:698 L:699 L:700 |
| 135    | 973.2                  | 1230                     | Blue   | W:27 C:28 T:29 I:30 E:34 F:60 A:61 L:62 V:76 T:77 L:78 <b>D:79</b> M:82 E:85 E:99 G:137 L:138 G:139 R:140 C:179 C:202 S:203 S:204 E:206 P:207 F:209 G:210 <b>Y:211</b> S:212 G:213 F:215 K:216 N:236 <b>H:272</b> G:313 Q:314 R:315 D:316 L:317 L:318 F:319 K:320   |
| 134    | 732.7                  | 1045.5                   | Cyan   | W:366 C:367 A:3678 V:369 E:372 E:373 L:411 N:412 L:413 <b>D:414</b> E:432 R:482 <b>D:527</b> <b>D:528</b> N:537 S:538 K:539 E:540 K:541 <b>Y:543</b> G:544 <b>Y:545</b> T:546 G:547 R:560 <b>H:614</b> T:655 K:656 N:657 L:658 L:659 F:670 N:671  |
| 133    | 605.7                  | 1183.9                   | Yellow | T:109 <b>H:110</b> A:567 Q:568 V:569 P:570 S:571 A:336 <b>Y:338</b> L:339 G:340 S:341 R:342 <b>Y:343</b> L:344 L:404 C:424 G:425 L:426 R:619 S:620 <b>D:621</b> C:703 L:706 T:707 R:708   |